Załącznik 2 do zarządzenia nr 16 Rektora Uniwersytetu Śląskiego w Katowicach
z dnia 28 stycznia 2015 r.

UNIWERSYTET SLASKI W KATOWICACH
WYDZIAL MATEMATYKI, FIZYKI I CHEMII

DIANA RIABUSHKINA

NR ALBUMU 298769

TECHNICZNA ANALIZA SZREGÓW CZASOWYCH INDEKSU DOW JONES

PRACA DYPLOMOWA LICENCJACKA

DOKTOR ŁUKASZ MACHURA. ZAKŁAD FIZYKI TEORETYCZNEJ UŚ
(INSTYTUT FIZYKI)

Słowa kluczowe:…………………………………

Oświadczenie autora pracy

Ja, niżej podpisana:
imię (imiona) i nazwisko: Diana Riabushkina
autor pracy dyplomowej pt .: „ANALIZA TECHNICZNA SZEREGÓW CZASOWYCH INDEKSU DOW JONES”
Numer albumu: 298769
Studentka Wydzialu Matematyki, Fizyki i Chemii Uniwersytetu Śląskiego w Katowicach
kierunku studiów: Ekonofizyka
specjalności: Ekonofizyka

Oświadczam, że ww. praca dyplomowa:

- została przygotowana przeze mnie samodzielnie1,
- nie narusza praw autorskich w rozumieniu ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (tekst jednolity Dz. U. z 2006 r. Nr 90, poz. 631, z późn. zm.) oraz dóbr osobistych chronionych prawem cywilnym,
- nie zawiera danych i informacji, które uzyskałem/łam w sposób niedozwolony,
- nie była podstawą nadania dyplomu uczelni wyższej lub tytułu zawodowego ani mnie, ani innej osobie.

Oświadczam również, że treść pracy dyplomowej zamieszczonej przeze mnie w Archiwum Prac Dyplomowych jest identyczna z treścią zawartą w wydrukowanej wersji pracy.

Jestem świadoma/y odpowiedzialności karnej za złożenie fałszywego oświadczenia.

Katowice, 17.02.2016 A

Data Podpis autora pracy

Spis treści

Wstęp

Każdego dnia zbieramy dane statystyczne z różnych dziedzin, które mogą dotyczyć odmiennych sfer. W ekonomii są to codzienne notowania giełdowe, kursy walut, prognozowanie wielkości sprzedaży lub popytu. Nie są to jedyne możliwe przykłady danych poddawanych obróbce statystycznej. Ponieważ warunki ekonomiczne zmieniają się z czasem to chcemy przewidzieć wpływ, jaki te przemiany będą miały na naszą firmę. Różny jest zakres informacji wymaganych w procesach decyzyjnych. Rozwój technologii informacyjno – komunikacyjnych sprawia, że przedsiębiorstwa gromadzą w formie elektronicznej wiele danych dotyczących działalności ich samych oraz ich klientów [1]. Informacje te najczęściej są indeksowane czasem, więc tworzą zbiory danych (ciągi danych) nazywane **szeregami czasowymi.** Dane poddane analizie mogą być bogatym źródłem informacji. Dzięki takiej analizie łatwiej jest podejmować decyzje obarczone mniejszym ryzykiem. Dla pojedynczych procesów ekonomicznych mamy do czynienia ze zmiennymi losowymi szokowymi, których realizacje obserwujemy w odstępach rocznych, kwartalnych itd. Nie można ponownie zaobserwować realizacji tej samej zmiennej ekonomicznej. Na tym polega trudność wnioskowania o zachowaniach podmiotów gospodarczych w dziedzinie czasu.

Spośród wszystkich spółek chciałabym przedstawić cztery: International Business Machines, Apple, Intel Corporation, Microsoft Corporation. Spółki te działają w branży IT, i znajdują się wewnątrz tego indeksu: notowane na giełdach NASDAQ i NYSE. Dlaczego wybrałam właśnie te spółki? Ponieważ są to duże przedsiębiorstwa informacyjne i elektroniczne, znane na całym swiecie. Zmiana cen akcji tych przedsiębiorstw ma duży wpływ na zachowanie indeksu Dow Jones.

Od czego zależy wartość akcji spółek? Czynnikiem determinującym jest pozycja firmy, jak również ogólny stan rynku. Żeby przewidzieć spadek akcji wprowadzono analizę na rynku akcji i indeksy giełdowe.

W rozdziale pierwszym omówiono historię indeksu Dow Jones, co to jest giełda NASDAQ , krótki opis wybranych spółek, wchodzących w skład indeksu. W rozdziale drugim wprowadzono podstawowe pojęcia związane z szeregiem czasowym. Omówimy metody dekompozycji, metodę różnicowania, modele ARIMA. Trzeci rozdział obejmuje część praktyczną, czyli analizę danych spółek Apple, IBM, Microsoft, Intel za pomocą środowiska R.

1.1 Historia indeksu Dow Jones

Na początku zadajmy sobie pytanie, czym jest indeks? Indeks daje nam krótką, systematyczną informację na temat rynku. Jest to pewna statystyczna miara zmian cen „portfela” pewnych akcji wybranych z całego rynku, ponieważ trudno policzyć pojedyncze spółki i na ich podstawie wyciągać ogólne wnioski. W sytuacji idealnej indeks powinien być tak skonstruowany, aby zmiana jego wartości dokładnie proporcjonalnie odzwierciadlała zmianę cen akcji objętych indeksem.

Indeks Dow Jones to jeden z największych światowych indeksów giełdowych, który jest obliczany w Stanach Zjednoczonych na podstawie 30 największych firm i odzwierciedla stan amerykańskiego przemysłu. Lista tych firm jest dynamiczna i może ulegać zmianie. Jest to pierwszy indeks giełdowy wynaleziony w USA, w II połowie XIX wieku. Założycielem spółki Dow Jones & Company jest Charles Dow oraz statystycy Edward Jones i Charles Bergstresser.

Pierwsze indeksy zaczęto drukować w gazecie „Wall Street Journal” w 1889 roku. Były to rewolucyjne publikacje, do tamtej pory nikt nie publikował raportów giełdowych. Prawo, które zobowiązywało spółki emitujące papiery wartościowe do publikowania raportów, zostało przyjęte w Stanach Zjednoczonych dopiero w 1934 roku. W drugim stuleciu istnienia WSJ jest liderem w grupie czasopism finansowych w Stanach Zjednoczonych oraz na całym świecie.

Charles Dow postanowił stworzyć na giełdzie tak zwany „barometr”, wskaźnik który będzie w stanie obliczać i wyrażać w postaci jednego pomiaru „samopoczucie” rynku w całości. Problem jest bardzo poważny ponieważ w tym samym czasie wartość akcji niektórych spółek rośnie, innych - maleje, trzecich – pozostaje bez zmian. Jak ocenić stan grupy spółek, a jak cały amerykański rynek w całości? Charles Dow znalazł odpowiedź na to pytanie. 3 lipca 1884 roku stworzył swój indeks. Metoda była prosta: wybrano 10 największych firm, których akcje są notowane na giełdzie i zaczęto codziennie wyliczać średnią wartość papierów tych firm. Niektóre akcje mogły iść w górę, inne spadać ale zmiana w ich średnich pozwalała zobaczyć ogólną tendencje. Był to pierwszy indeks Dow Jones. Po raz pierwszy został on oficjalnie opublikowany w 1896 roku. Wzrost indeksu był związany ze wzrostem wartości tych firm, a spadek był związany ze spadkiem aktywności gospodarczej i zmniejszeniem kapitału głównych podmiotów gospodarczych. Ciekawe jest to, że spośród dziesięciu największych firm wybranych przez Charlsa Dow'a i Edwarda Jonesa dziewięć było przedsiębiorstwami kolejowymi. Do 1928 roku ilość spółek, wchodzących do grupy obliczanych przy rozrachunku indeksu Dow Jonesa zwiększyła się do trzydziestu. Interesującym jest to, że ilość spółek od tego czasu nie uległa zmianie.

Podstawą dla obliczenia indeksu Dow Jones są tylko akcje notowane na New York Stock Exchange, wybór nie zależy od pozycji i sławy firmy. Przykładowo, Microsoft nie ma żadnego wpływu na kształtowanie indeksu, ponieważ jego akcje są notowane na giełdzie NASDAQ. Obliczany jest również indeks NASDAQ, który charakteryzuje aktywność transakcji off-walut kupna i sprzedaży papierów wartościowych. Obliczanie indeksu Dow Jonesa w naszych czasach jest bardzo skomplikowane, a średnia arytmetyczna początkowo będąca podstawą indeksu, nie ma już znaczenia.

Skoro tak, to dlaczego indeks Dow Jonesa jest tak popularny? I czy podstawowa ilość firm może mieć znaczący wpływ na jego wartość? Najprawdopodobniej nie. Dzisiaj według ekspertów zmiana indeksu zależy od bezwzględnej ceny akcji. Jednak indeks Dow Jones nie traci swojej popularności.

1.2 Czym jest NASDAQ?

Sprzedaż akcji dużych firm w Stanach Zjednoczonych odbywa się na rynkach aukcyjnych. Największym takim rynkiem jest New York Stock Exchange (NYSE), na którym notowane jest ponad 85% wszystkich akcji. Duże znaczenie w gospodarce amerykańskiej ma regulowany rynek pozagiełdowy. W 1971 roku Krajowe Stowarzyszenie Dealerów Papierów Wartościowych (National Association of Securities Dealers, NASD) udostępniło dealerom i brokerom elektroniczny system nazywany NASDAQ (Automatyczny System Notowań NASD).

Indeks NASDAQ (National Association of Securities Dealers Automated Quotations system) – indeks rynku OTC (Over-the-Counter), który jest skomputeryzowanym systemem obrotu pomiędzy dealerami papierów wartościowych, którzy podają okresowe notowania cen papierów wartościowych. Dealerzy ci są kreatorami rynku papierów wartościowych notowanych przez organizację NASDAQ, składają oni oferty cenowe kupna i sprzedaży, które wyrażają ich gotowość do realizacji określonych zleceń.

NASDAQ Stock Exchange została założona 8 lutego 1971 roku, a pierwsza aukcja odbywała się przez telefon. W 1987 roku, po załamaniu się rynku finansowego, kierownictwo amerykańskiej giełdy NASDAQ pozwoliło na składanie ofert elektronicznych. Obecnie na giełdzie NASDAQ notuje się ponad 3200 firm i około 5000 papierów wartościowych. Patrząc na ilość firm i wolumenu akcji dziennie NASDAQ jest największą giełdą w Stanach Zjednoczonych.

Notowania NASDAQ są dostępne na trzech poziomach:

I poziom - przekazywanie informacji na temat aktualnych notowań cenowych;

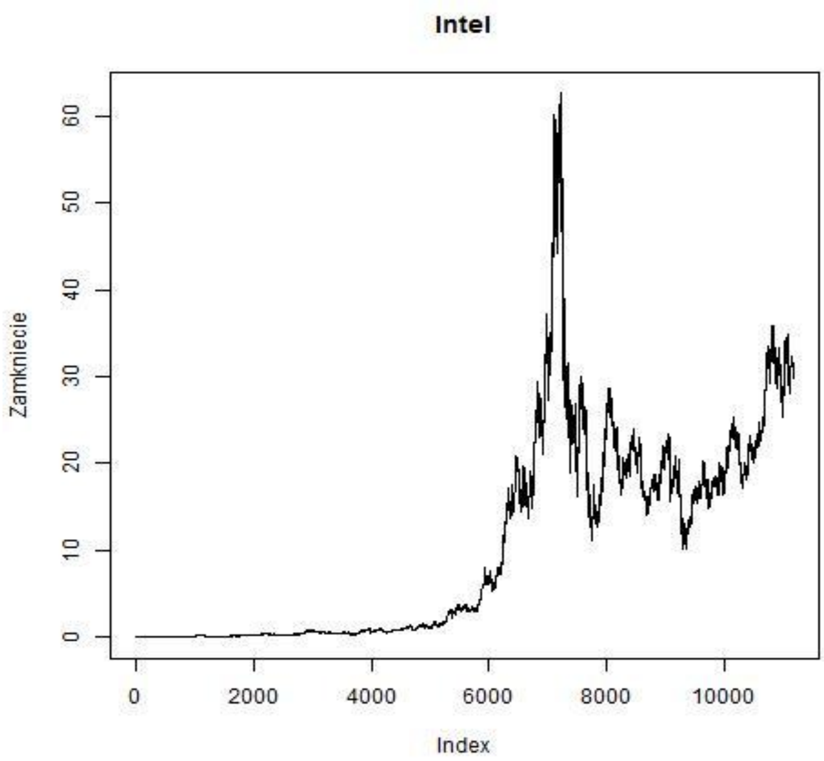
II poziom - terminal łączy osoby tworzące rynek z brokerami i innymi dealerami oraz pozwała na przegląd notowań cenowych proponowanych przez wszystkich kreatorów rynku działających w tej sieci; terminale tego poziomu umożliwiają także dostęp do kwotowań wewnętrznych, czyli do najwyższych notowań ofert cenowych kupna i najniższych notowań ofert sprzedaży papierów wartościowych.

Na **III poziomie** terminale pozostają tylko do dyspozycji kreatorów rynku. Pozwalają one dealerom sieci NASDAQ wprowadzać lub zmieniać informacje dotyczące podawanych cen notowań.

Na chwilę obecną w systemie NASDAQ notowanych jest wiele innych amerykańskich i zagranicznych firm, więcej w porównaniu ze wszystkimi giełdami amerykańskimi. Również system techniczny pozwala NASDAQ umieścić w nim wszystkie spółki które są obecne na giełdach.

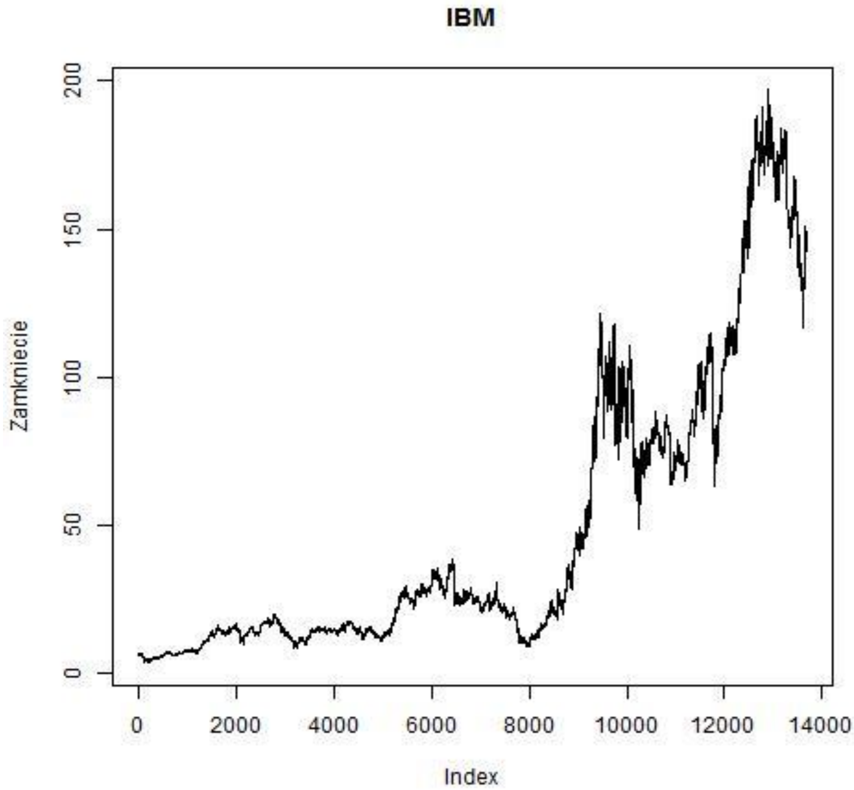
1.3 Wybrane spółki indeksu Dow Jones – opis, notowania giełdowe

Intel Corporation to amerykańska korporacja, która produkuję szeroką gamę urządzeń elektronicznych i detale dla komputerów, mikroprocesory, pamięć RAM i Flash, mikrokontrolery, a także urządzenia sieciowe. Siedziba znajduje się w mieście Santa Clara w stanie Kalifornia. Intel jest największym na świecie wytwórcą mikroprocesorów, obsługującym 75% tego rynku. Głównymi konsumentami produktów firmy są producenci komputerów DELL, Hawllet-Packard i Apple. Największymi konkurentami są produkty firm AMD, IBM, Cyrix i Motorola. Założycielami firmy byli Gordon E. Moore oraz Robert Noyce. Powstała ona 18 lipca 1968 roku, a nazwa pochodzi od słów „Integrated Electronics". Cena akcji spółki wynosi 32,78$ i nie ma dużego wpływy na ksztaltowanie indeksu DJIA. Wykres na rysunku 1.1 przedstawia oryginalny szereg czasowy zachowania notowań giełdowych (ceny zamknięcia) .



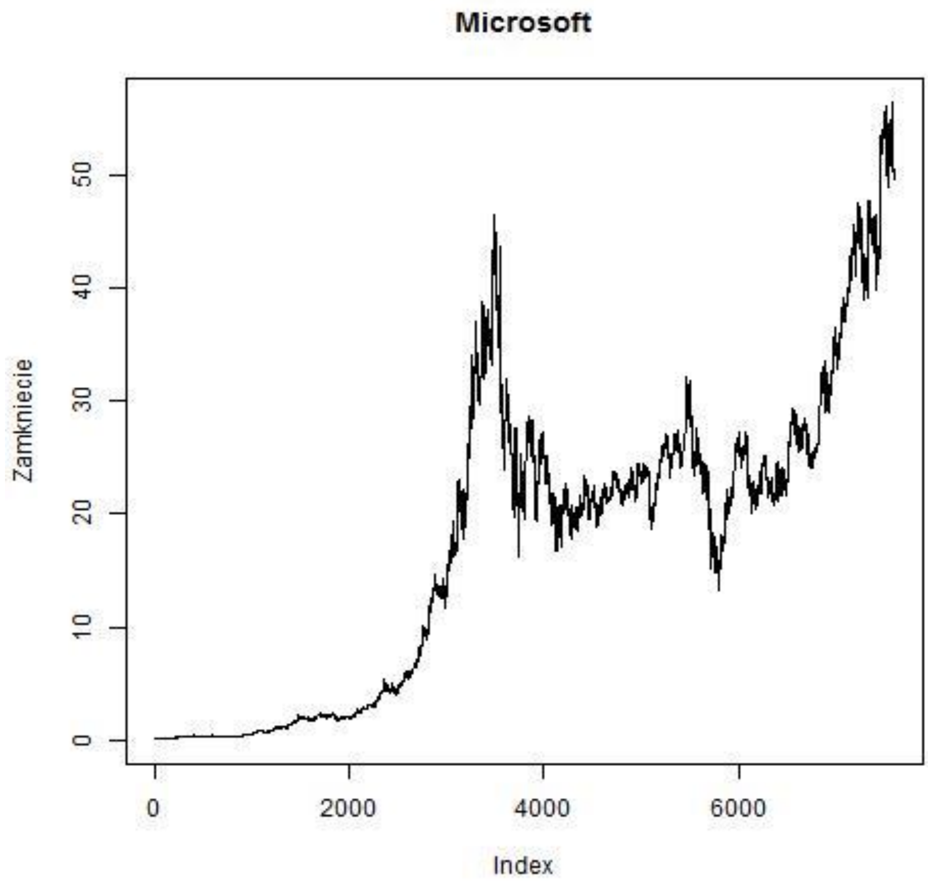
Rysunek 1.1. Codzienne notowanie INTEL (ceny zamknięcia) od 01.02. 1972 roku do 06.05.2016 roku (dane pobrane ze strony http://stooq.com/q/d/?s=intc.us)

Inrernational Business Machines Corp. jest międzynarodową korporacją z siedzibą w Armonk. Rozpoczęła swoją działalność jako The Computing-Tabulating Recording Company w 1911 roku, ale w 1924 roku zmieniła profil swojej działalności. Po raz pierwszy została włączona do indeksu w 1932 roku, ale została usunięta przez United Aircraft. IBM wrócił do indeksu w 1979 roku. Dziś to jeden z największych producentów i dostawców sprzętu komputerowego, oprogramowania, usług informatycznych i konsultingowych na świecie. Cena akcji wynosi 151,73$ i ma duży wpływ na ksztaltowanie indeksu. Wykres na rysunku 1.2 przedstawia szereg czasowy oryginalnych danych zachowania notowań giełdowych (ceny zamknięcia) dannej spólki.



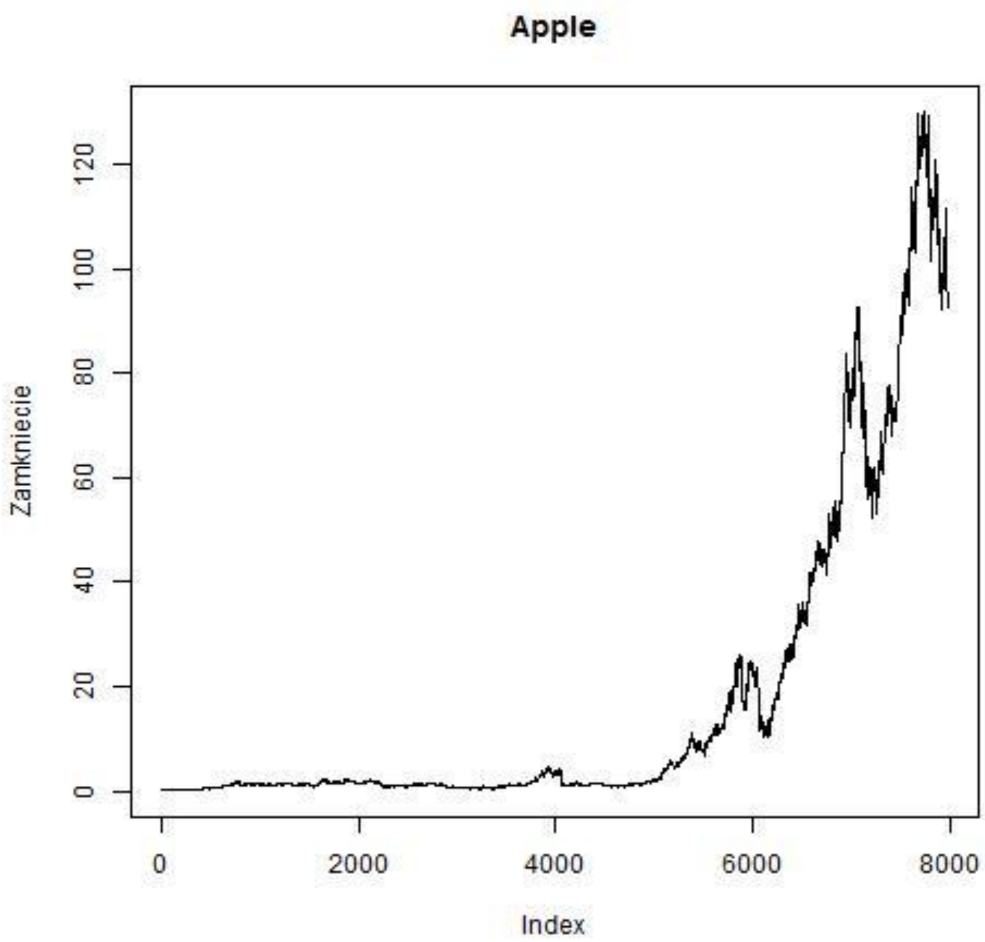
Rys.1.2. Codzienne notowanie IBM (ceny zamknięcia) od 02.01. 1962 roku do 06.05.2016 roku
(dane pobrane ze strony http://stooq.com/q/d/?s=ibm.us)

Microsoft Corporation - ten gigant zaczął swoją działalność w kwietniu 1975 roku jako projektant oprogramowania komputerowego, którym zdominował branżę. Przesądziło o tym wypuszczenie pierwszej wersji MS – DOS w 1981 roku. Po czterech latach na rynku ukazał się Windows 1.0, oferując pierwszą wersję swojego graficznego interfejsu użytkownika (GUI) dla użytkowników DOS. Założycielami spółki są Bill Gates i Paul Allen. Cena akcji wynosi 50,53$, i nie ma żadnego wpływy na ksztaltowanie indeksu, ponieważ jest notowany na innej giełdzie. Wykres na rysunku 1.3 przedstawia szereg czasowy zachowania notowań giełdowych (ceny zamknięcia) spółki Microsoft.



Rys.1.3. Codzienne notowanie Microsoft (ceny zamknięcia) od 13.03. 1986 roku do 06.05.2016 roku (dane pobrane ze strony http://stooq.com/q/d/?s=mcst.us)

Historia firmy **Apple** zaczęła się 30 lat temu, kiedy dwóch kolegów - Steve Jobs i Steve Wozniak postanowili założyć własną firmę, aby produkować i sprzedawać komputery. Oficjalnie Apple powstała 1 kwietnia 1976 roku. Korporacja Apple Inc. weszła do składu indeksu Dow Jones Indastrial Avarege 18 marca 2016 roku zamiast telekomunikacyjnej kompanii AT&T. Kompania weszła do indeksu pod warunkiem podziłu akcji w stosunku 7:1. 6 czerwca 2016 był ostatni ostatnim dniem, w którym notowane były stare akcje. W wyniku tego podziłu każdy akcjonariusz Apple, za jaedną starą akcję otrzymał siedem nowych, każdą z nich sieden razy mniej wartą. „Jako największa korporacja w świecie i lider branży technologii, Apple jest oczywistym wyborem dla DJIA”- oświadczył David Blitzer, przewodniczący indeksowego komitetu S&P Dow Jones Indices. Apple Inc. to najbogatsze przedsiębiorstwo w świecie. Na początku lutego jej kapitalizacja przekroczyła $700 mld, a do 23 lutego osiągnęła $775mld. Obecnie akcje Apple są poziomie $126.6. Wykres na rysunku 1.4 przedstawia szereg czaasowy zachowania notowań giełdowych (ceny zamknięcia) spółki Apple.



Rys.1.4 Codzienne notowanie spółki APPLE (ceny zamknięcia) od 07.09. 1986 roku do 06.05.2016 roku (dane pobrane ze strony http://stooq.com/q/d/?s=appl.us)

2.1 Definicja i cele szeregów czasowych

Szereg czasowy jest zbiorem wartości badanej cechy, opisującym zjawiska zaobserwowane w różnym czasie, może również stanowić ciąg obserwacji x_t, zapisywanych w ściśle określonym czasie.

Składowe części szeregu czasowego to składniki systematyczne oraz przypadkowe, zwane inaczej składnikami losowymi lub wahaniami przypadkowymi. Procesy deterministyczne określa składowa systematyczna, a składowa losowa – proces stochastyczny. W modelach deterministycznych opis systemu i otoczenia nie zawiera zmiennych losowych; proces stochastyczny to z kolei zbiór zmiennych losowych określonych na przestrzeni probabilistycznej S, które przyjmują wartości w przestrzeni mierzalnej M.

Wśród składowych systematycznych szeregu czasowego możemy wyróżnić:

- Trend (tendencję rozwojową),
- Wahania sezonowe,
- Wahania cykliczne (koniunkturalne),
- Wahania przypadkowe

Obserwacje występujące w szeregach czasowych mogą różnić od pozostałych obserwacji w szeregu. Są to wartości nietypowe dla danego zjawiska, spowodowane różnego rodzaju zaburzeniami w jego przebiegu, dlatego trzeba zwrócić uwagę na przyczyny występowania wartości odstających w szeregu, ponieważ mogą one obrazować ważne wydarzenia ekonomiczne, mające znaczący wpływ w przebiegu całego szeregu czasowego.

Podstawowe cele badania szeregów czasowych:

1. wykrywanie natury danego zjawiska losowego, tj. badanie własności szeregu i znalezienie modelu najlepiej opisującego zjawisko,
2. prognozowanie (predykcja), tj. przewidywanie kolejnych wartości szeregu czasowego na podstawie znalezionego modelu.

Oba te cele wymagają rozpoznawania i opisywania, w sposób formalny, elementów szeregu czasowego. Analizując szereg czasowy, w pierwszej kolejności wybieramy właściwy model (klasę modeli) reprezentujący dane. Każda obserwacja x_t jest konkretną realizacją jakiejś zmiennej losowej X_t. Szereg czasowy $\{x_t, t \in T_0\}$ będzie zatem realizacją całego zbioru zmiennych losowych $\{X_t, t \in T_0\}$. Na podstawie realizacji zmiennej w okresie t można coś powiedzić o realizacji zmiennej w okresie t+s. Dla dowolnych momentów s i t możemy wyznaczyć funkcje:

1. średniej (wartość oczekiwaną) : $\mu_t = E(X_t)$
2. autokowariancji

Dla dwóch zmiennych losowych $\{X_t, t\epsilon\ T\}$ oraz $\{Y_s, s\ \epsilon\ T\}$ funkcja

cov (X(t), Y(r)) = E[$(X_t - EX_t)(Y_s - EY_s)$] =$E(X_t, Y_s) - EX_t EY_s$

dla t,s ϵ T określą liniową zależność pomiędzy powyższymi zmiennymi losowymi. Stopień współzależności owych zmiennych losowych można podać za pomocą współczynnika korelacji Pearsona r_{XY} cov(X,Y) = $r_{XY}\sigma_X\sigma_Y$. Wspólczynnik korelacji mówi nam o sile związku i mieści się w przedziale domkniętym [-1,1]. Im większy jest współczynnik, tym silniejsza jest zależność zmiennych losowych między zmiennymi. r_{XY}=0 oznacza brak liniowej zależnośći między cechami, gdy $r_{XY} = -1$ to można powiedzić, że gdy wzrastają wartości jednej zmiennej to maleją wartości drygiej, $r_{XY} = 1$ - odwrotnie.

3. autokorelacji

Dla szeregu X_t funkcja autokorelacji ACF (Autocorrelation Coefficient Funktion) rzędu q jest zdefiniowana następująco : $ACF(q) = \frac{cov(X_t, X_{t-q})}{var(x_t)} = \frac{\sum_{i=1}^{n}(x_i-\bar{x})*(y_{i-q}-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}$ dla q=0,1,2...

gdzie $\bar{x}$ to średnia wartość x, $\bar{y}$ to średnia wartość y, q to przesunięcie.

Funkcja ACF jest miarą zależności (korelacji liniowej) pomiędzy obserwacjami szeregu czasowego odległymi o q jednostek czasowych. Parametr q jest tak zwanym parametrem opóżnienia. Za pomocą wykresu funkcji ACF (korelogram) łatwo identyfikować podstawowe własności analizowanego szeregu. Funkcja ACF, zanikająca powoli i cyklicznie, wskazuje na obecność sezonowości. Szybkość zanikania ACF identyfikuje rząd dla modeli MA.

Funkcja PACF(q) mierzy odległość pomiędzy obserwacjami odległymi o q jednostek czasowych $X_t\ a\ X_{t+q}$. Jeżeli wartość opóżnienia PACF jest bliska 1 to może świadzyć o obecności silnego trendu. Funkcja autokorelacji cząstkowej PACF jest użyteczna do identyfikacji rządu modeli AR(p).

2.2 Stacjonarność szeregów

Przed właściwym modelowaniem koniecznie jest odpowiednie przygotowanie danych, czyli wykonanie przekształceń które mogą ułatwić dopasowanie modelu. Duże znaczenie ma identyfikacja regularności występowanie trendu oraz wahań sezonowych w analizowanym szeregu co jest oznaką szeregów niestacjonarnych.

Proces generujący dane (stochastyczny) $\{X_t \in Z\},\ gdzie\ Z = \{0, \pm1, \pm2 \dots\}$ nazywamy słabo stacjonarnym, jeżeli można określić następujące jego parametry:

1. wariancję - E$|X_t| < \infty\ dla\ t \in Z$,
2. wartość oczekiwaną - E$X_t = m\ dla\ t \in Z$,
3. autokowariancję, która zależy jedynie od rzędu opóżnienia, czyli od tego, na ile odlegle od siebie w czasie są obserwacjy – $\gamma_X(r, s) = \gamma_X(r + t, s + t) dla\ t \in Z$.

Szeregami niestacjonarnymi mogą być szeregi cen różnych dóbr i instrumentów finansowych. Wariancja szeregu może rosnąć w okresie bieżącym pod wpływem zmian z okresów poprzednich, co powoduje, że pojedyncze skoki mogą mieć długotrwały wpływ na kształtowanie się szeregu. Takie własności mogą wykazywać szeregi stóp wzrotu z inwestycji w akcje lub waluty. Dla ich kształtowania się ważną rolę odgrywają nastroje i oczekiwania inwestorów.

Proces stochastyczny nazywamy ściśle stacjonarnym, jezeli łączny rozkład zmiennych losowych X(t) oraz X(t+ϵ) jest identyczny

$$f(x_1; t_1; x_2; t_2 \dots; x_n; t_n) = f(x_1; t_{1+\epsilon}; x_2; t_{2+\epsilon}; \dots.; x_n; t_{n+\epsilon}).$$

$f(x_1; t_1) = f(x_1; t_{1+\epsilon})$ *oznacza*, że funkcja rozkładu nie zależy od czasu. Wynika z tego natychmiast, że wartość oczekiwana jest również stałą funkcją czasu $\langle m(t) \rangle = m.$ Szeregi stacjonarne nie powinny wykazywać żadnych trwałych tendencji zmian, dlatego łatwiej przewidywać ich przebieg. Wśród zmiennych ekonomicznych szeregi stacjonarne występują w mniejszości. Z tych powodów musimy nauczyć się sprowadzać szeregi niestacjonarne do postaci stacjonarnej. Z analizy wynika, że niestacjonarność występuje ze względu na średnią lub wariancję. Jeżeli znajdziemy takie trwałe tendencje zmian w szeregu czasowym, to możmy sprawdzić, czy stopień odchylenia badanego szeregu lub jego zmienności od trendu (w średniej lub warianscj) są stacjonarne. Rozumowanie takie leży u podstaw różnych metod eliminacji trendu z szeregu czasowego. Metoda filtracji (wygładzanie) szeregu jest mniej popularnym sposobem, zmierzającym do tego, aby wyróżnić w nim część systematyczną. Szereg stacjonarny w sensie słabym możemy otrzymać w wyniku usunięcia trendu i dalej modelować za pomocą metod dla szeregów stacjonarnych. Modele liniowe możemy badać za pomocą modeli ARMA (autoregresyjne modele średniej kroczącej).

Szeregi, w których nie występują trendy, sezonowość oraz efekty cykliczne nazywamy szeregami niestacjonarnymi, a te, w których zmienność (wariancja) jest stała w czasie - szeregami stacjonarnymi.

Najbardziej popularne modele stacjonarne:

- WN(0,σ^2) biały szum o średniej zero i wariancji σ^2(White Noise),
- AR(p) model autoregresji rzędu p (AutoRegressive model of order p),
- MA(q) model średniej krączącej rzędu q (Moving Avarage model of order q).
- ARMA(p,q): model mieszany AR(p) i MA(q).

2.3 Przekształcenie szeregów czasowych metodą różnicowania

Przed przystąpieniem do analizy musimy przygotować dane. Stosując przekształcenia możemy wyeliminować źródła zmienności. Dzięki temu model staje się łatwiejszy do interpretacji i pozwala uzyskać lepsze prognozy. Specyfika danych może wymagać przekształcenia wstępnego. Do danych najczęściej dopasowują się dane stacjonarne, które wymagają, aby nie występowały regularne tendencje długoterminowe oraz sezonowość. Wykorzystując transformacje możemy przekształcić nasz szereg tak, aby spełniał te założenia.

Jedną z najczęściej stosowanych metod przeksztalsenia szeregów czasowych jest różnicowanie (ang. differencing) stosowana dla usunięcia z danych trendu. Różnicowanie stosujemy również do sprowadzenia danych do postaci stacjonarnej. Szereg właściwości którego nie zmieniają się w czasie nazywamy stacjonarnym. Szeregi w których występują sezonowość lub trend długoterminowy nazywamy szeregami niestacjonarnymi.

Różnicowanie sprowadza się do zastąpienia oryginalnych danych szeregiem różnic (przyrostów), wyznaczonych dla ustalonych opóźnień czasowych. Żeby wyeliminować z danych trend liniowy, kwadratowy, a także sezonowość dobieramy odpowiednio parametry różnicowania. Modele ARIMA jest ściśle związane z operacją różnicowania. Różnicowanie z opóźnieniem 1 jest podstawową wersją operacji różnicowania. Różnicowanie polega na tym że na podstawie oryginalnych danych $X_1, X_2 \dots, X_n$ wyznaczamy szereg różnic (przyrostów) odejmując wartości w następujących po sobie chwilach $X_2 - X_1, X_3 - X_2 \dots . . X_n - X_{n-t}$ i otrzymujemy w ten sposób nowy szereg $Y_t = X_t - X_{t-1}$, gdzie t= 2,3....n.

Definicja (Różnicowanie z opóźnieniem 1)

Operator różnicowania z opóźnieniem jeden definiuemy jako

$$\nabla X_t = X_t - X_{t-1}.$$

Zróżnicowany szereg również możemy przedstawić natępująco

$$\nabla X_t = X_t - X_{t-1} = (1 - \beta) X_t,$$

gdzie β jest operatorem przesunięcia wstecz (backward shift operator) zdefiniowany jako

$$B X_t = X_{t-1}.$$

Czasami jednokrotne roznicowanie nie jest wystarczające żeby usunąć w szeregu trend wtedy można zastosować wielokrotne różnicowanie. Jednokrotne różnicowanie nie wyeliminuje sezonowości i może być stosowane do szeregów gdzie występuje tylko trend liniowy. Aby pozbyć się sezonowości potrzebujemy użyć operacji róznicowania z opóźnieniem sezonowym s.

2.4 Metody dekompozycji szeregów czasowych

W szeregach czasowych możemy zaobserwować bardzo zróżnicowane zachowanie, dlatego zaczynamy analizę dowolnego szeregu od wizualizacji danych. Na początku powinniśmy stworzyć wykres badanych danych. Daje nam to możliwość oszacowania, jak konkretny zbiór danych zachowuje się w czasie. Zazwyczaj, oprócz losowych fluktuacji, często występuje trend oraz sezonowość. W trudniejszych przypadkach należy zadecydować, czy nie podzielić danego szeregu na dwie czy więcej części i badać je osobno.

Dzięki dekompozycji możemy wyeliminować sezonowość z szeregu czasowego. Wykonanie dekompozycji może ułatwić znalezienie adekwatnego modelu dla szeregu czasowego. Dekompozycję przeprowadzamy po to, aby określić tendencję rozwojową, stwierdzić obecność lub brak trendu i sezonowości, określić ich charakter. Jeżeli występują trzeba usunąć trend, wyeliminować sezonowość i zastosować metody prognozowania dla szeregów stacjonarnych (modele AR, MA, ARMA), które pozwalają na wyznaczenie prognoz dla losowych fluktuacji (reszt).

Dekompozycję szeregu czasowego można opisać formułą

$$X_t = f(s_t, m_t, Z_t)$$

gdzie: f()- pewna funkcja, $s_t - składowa\ sezonowa$, $m_t -$ trend długoterminowy, $Z_t -$ zakłócenie losowe.

Wszystkie trzy składowe nie zawsze występują w każdym szeregu czasowym. Najczęściej może występować tylko trend ($X_t = m_t + Y_t$) lub sezonowość ($X_t = s_t + Y_t$).

W praktyce wśród modeli dekompozycji wyróżniją dwa to:

dekompozycja addytywna

$$(X_t = s_t + m_t + Y_t)$$

dekompozycja multiplikatywna

$$(X_t = s_t \cdot m_t \cdot Y_t)$$

Wybór modeli zależy od badanego szeregu.

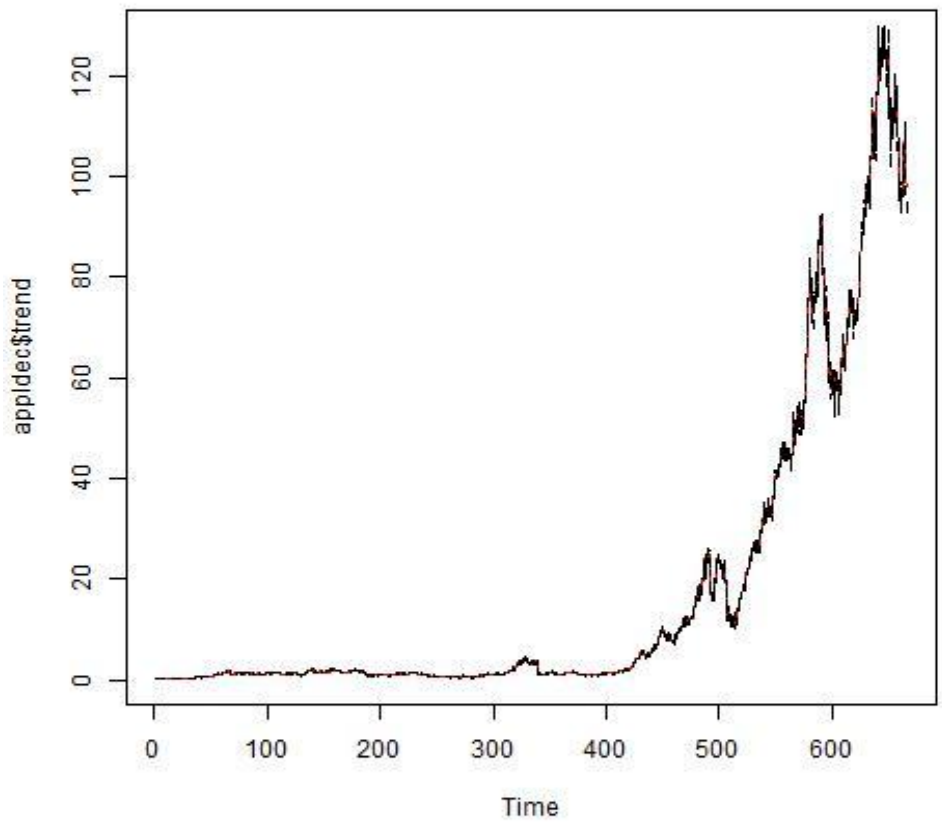
Dla dekompozycji szeregu stosowane są różne metody. Wśród dostępnych metod możemy wydzielić dwie grupy:

- Metody parametryczne – przyjmujemy określony model, dekompozycja sprowadza się do znaleziena nieznanych estymatorów.
- Metody nieparametryczne – nie otrzymujemy trendu w postaci wzoru analitycznego, bo nie ograniczamy się do modeli z określonej klasy.

2.4.1 Wygładzanie za pomocą MA (Moving Avarage)

W wypadku szeregu czasowego o bardzo niewyraźnym trendzie i wahaniach sezonowych również dla krótkich szeregów czasowych ma zastosowanie jedna z najbardziej popularnych metod wygładzania wykładniczego. Jej idea polega na tym że wygładzamy szereg czasowy poprzez uśrednienie sąsiadujących wartości, aby wyeliminować losowe fluktuacje. Istatną rolę w przypadku wygładzania danych metody średniej kroczącej odgrywa odpowiedni dobór rzędu

q. Mały rząd q powoduje dużą zmienność i niewystarczające wygładzanie, a zbyt duży rząd q charakterezuje zbyt duże wygładzanie.



Rys.2.3 Notowania giełdowe spółki Apple metodą wygładzania średniej ruchomej

Proces średniej kroczące MA jest to model liniowy w którym jeżeli $\varphi(z) = 1\ to\ X_t = \theta(B)Z_t$ i proces X_t nazywamy procesem średniej kroczącej rzędu $q \equiv MA(q)$. Rozwiązaniem równania $X_t = \theta(B)Z_t$ jest proces stacjonarny. Proces MA, nazywamy odwracalnym, jeżeli da się sprowadzić go do stacjonarnego procesu autoregrysyjnego. Warunkiem odwracalności jest $\theta_1 < 1$.

Żeby dokonać dekompozycji powinniśmy oszacować składnik trendu m_t, wykorzystując metodę symetrycznej średniej kroczącej. Potem oszacowany trend usuwamy z danych. Wyznaczmy szereg $X_t - m_t$ w przypadku dekompozycji addytywnej. W ostatnim kroku wyznaczmy losowe zakłócenia (reszty), usuwając trend i sezonowość. W przypadku dekompozycji addytywnej $Y_t = X_t - m_t$.

Model MA(q) uwzględnia obecność korelacji czasowej wystpującej w szeregu czasowym. Rząd q i współczynniki medelu $\theta_1, \dots, \theta_n$ decedują o tym, jak silna będzie korelacja i jaki ona będzie miała charakter

2.5 Model Atoregresji AR

Procesem autoregresyji p AR(p) jest nazywany proces w postaci:

$$Y_t = \sum_{i=1}^{p} \phi X_{t-i} + Z_t \text{ gdzie:}$$

$\phi_i - parametry\ modelu,$

$Z_t \sim N(0, \sigma_\varepsilon) - biały\ szum.$

Modelem autoreresji rzędu p nazywamy stacjonarny szereg czasowy X_t spelniający równanie

$$X_t = \phi X_{t-1} + \dots + \phi_p X_{t-p} + Z_t,$$

Gdzie $Z_t - szereg\ typu\ biały\ szum\ \left(WN(\sigma^2)\right).$

Wartość szeregu czasowego w chwili t jest przedstawiona jako liniowa kombinacja p wczesniejszych wartości $X_{t-1} \dots, X_{t-p}\ do$ której dodajemy zaklócenia w postaci białego szumu.[2]

Dla procesów AR(p) funkcja autokorelacji ACF maleje wykładniczo lub jest sinusoidą o zmniejszanej się amplitudzie wachań, funkcja PACF przyjmuje wartości równe zero dla opóżnień większych od p.

Żeby dobrze odzwierciedlić korelację czasową w szeregu analizowanym, modele średniej kroczącej MA(q) oraz modele autoregresji AR(p) wymagają dość często dużych rzędów q lub p odpowiednio. Może to spowodować problemy na etapie dopasowania modelu (mamy dużą liczbę współczynników). Wtedy alternatywą może być dopasowywanie modelu mieszanego ARMA, który łączy oba modela AR i MA, które wymagają mniejszej liczny parametrów, aby uzyskać dobre dopasowywanie do danych.

2.6 Modele ARMA

Modelem autoregresji średniej kroczącej ARMA (q,p) nazywamy stacjonarny szereg czasowy spełniający równanie

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q},$$

gdzie Z_t – szereg typu biały szum (WN(σ^2))

Model ARMA można zapisać w bardziej zawartej formie

$$\phi(B)X_t = \theta(B)Z_t, \text{ t=1,±1,±2,....}$$

Gdzie $\phi\ oraz\ \theta\ to$ woelomiany odpowiednio p-tego i q-tego stopnia.

$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p$

$\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$

nie mają wspólnych czynników (tzn. nie mogą być zredukowane do wielomianów mniejszych stopni).[]

Wygodnie jest przedstawić model ARMA(p,q), wykorzystując równoważny zapis operatorowy

$$\phi(B)X_t = \theta(B)Z_t$$

gdzie B- operator przesunięcia wstecz (backward shift operator) (to znaczy $BX_t = X_{t-1}$), $\phi(\cdot) -$ wielomian autoregresyjny stopnia p, $\theta(\cdot)$- wielomian średniej ruchomej stopnia q.

Modele AR(p) i MA(q) są szczególnymi przypadkami modelu ARMA(p,q). Rodzina modeli ARMA to uogólnienie zdefiniowanych wcześniej modeli AR(p) i MA(q). Proces ARMA (p,q) jest stosowany wyłącznie do szeregów stacjonarnych. Szereg ARMA jest stacjonarny, jeżeli spełnia założenie

$$\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0 \; dla \; |z| = 1$$

(wielomian autoregresyjny $\phi(z)$ nie ma pierwiastków na okręgu jednostkowym).

Poza stacjonarnością należy przyjąć także inne założenia, które gwarantują kolejne własności modelu ARMA. Wśród nich można wyróżnić:

- przyczynowość (causality) - możliwość przedstawienia modelu ARMA jego szereg MA(∞). Przyczynowość oznacza zatem możliwość przedstawienia wartości szeregu w chwili t jako liniowej kombinacji bieżącego i wcześniejszych zakłóceń.
- odwracalność (invertibility) – możliwość przedstawienia modelu ARMA jako szereg AR(∞). Odwracalność oznacza zatem możliwość przedstawienia wartości szeregu w chwili t jako liniowej kombinacji wcześniejszych wartości.

Aby model ARMA był przyczynowy i odwracalny, potrzebne są dodatkowe ograniczenia na współczynniki $\{\phi_i\}$ i $\{\theta_i\}$.

Przyczynowość jest równoważna warunkowi $\phi(z) = 1 - \phi_1 z - \cdots - \phi_p z^p \neq 0$

Z kolei odwracalność jest równoważna założeniu, że $\theta(z) = 1 + \theta_1 z + \cdots + \theta_q z^q \neq 0$

Najsilniejsza zależność wielokrotności pewnego opóźnienia sezonowego s często zależy od wcześniejszych obserwacji. Na przykład, dla danych miesięcznych może występować silna zależność okresowa dla opóźnień będących wielokrotnościami okresu s=12.

Dla procesów ARMA (p,q) funkcja autokorelacji ACF oraz funkcja autokorelacji cząstkowej PACF zanikają w taki sposób:

funkcja autokorelacji ACF po q -p odstępach czasu jest sumą zanikających funkcji wykładniczych i tłumionych sinusoid, funkcja autokorelacji cząstkowej PACF po q -p odstępach czasu jest zdominowana przez zanikające funkcje wykładniczych i tłumionych sinusoid.[2]

2.7 Model niestacjonarny ARIMA

Modele klasy ARMA (w tym modele AR(p), MA(q), ARMA(p, q) które rozpatrywaliśmy wcześniej mogą być zastosowane w przypadku, gdy analizowane dane można uznać za realizacje szeregu stacjonarnego. Oznacza to: brak trendów, jednorodną wariancję, a także szybko zanikającą (wraz z opóźnieniem czasowym h) funkcję autokorelacji ACF. Jeżeli analizowany szereg wykazuje odstępstwa od stacjonarności, możemy zastosować odpowiednie transformacje danych, które pozwolą przekształcić szereg do postaci stacjonarnej.

W celu usunięcia z szeregu niestacjonarności, co z reguły wiąże się z identyfikacją oraz usunięciem jego składowych deterministycznych (np. trend, sezonowość), można wykorzystać m.in. metody dekompozycji i różnicowanie. Można także uwzględnić składowe regularne bezpośrednio w postaci modelu. Dzięki temu możliwe jest uzyskanie modeli dla szeregów niestacjonarnych.

Spośród modeli niestacjonarnych, dotyczących wspomnianych metod eliminacji składowych regularnych możemy wyróżnić:

- model klasycznej dekompozycji $X_t = m_t + s_t + Y_t$
- modele ARIMA (AutoRegressive Integrated Moving Average) i SARIMA (Seasonal ARIMA)

Model ARIMA można określić jako model ARMA po wykonaniu operacji z opóźnieniem równym 1.

DEFINICJA

Szereg $\{X_t\}$ nazywamy procesem ARIMA(p,d,q), jeżeli po d-krotnym ($d \geq 0$) zróżnicowaniu z opóźnieniem 1 (zapisujemy: $Y_t = (1 - B)^d X_t$) jest już procesem ARMA(p,q), tzn X_t spelnia równanie

$$\phi^*(B)X_t \equiv \phi(B)(1 - B)^d X_t = \theta(B)Z_t, \;\; Z_t \sim WN(0, \sigma^2),$$

Gdzie $\phi(z) i \, \theta(z)$ są wielomianami odpowiednio stopnia p i q oraz $\phi(z) \neq 0 \; dla \; |z| \leq 1$

Należy zwrócić uwagę na to, że szereg Xt jest stacjonarny pod warunkiem, że d = 0. Modele ARIMA, które obejmują wiele szeregów niestacjonarnych, stanowią swego rodzaju generalizację modeli ARMA. Dzięki nim można modelować szeregi z trendem długoterminowym.

Korzystając z modelu ARIMA możliwe jest również modelowanie szeregów, których nie dotyczy trend deterministyczny. Jako jeden z przykładów należy podać model błądzenia losowego $X_t = X_{t-1} + Z_t$.

Niestety, model ARIMA nie pozwala uwzględnić sezonowości, która może wystąpić w analizowanym szeregu czasowym. W tej sytuacji bardziej adekwatnym będzie natomiast model SARIMA.

2.8 Identyfikacja modeli

Aby wybrać model odpowiedni do analizy szeregu czasowego należy zdecydować, który będzie najbardziej właściwy dla zgromadzonych danych. Można tego dokonać dzięki identyfikacji postaci modelu oraz wyborowi odpowiedniej złożoności modelu (tj. rzędów).

Należy pamiętać o tym, że reguły identyfikacji opierają się na własnościach teoretycznych (zwłaszcza charakterystycznych zachowaniach ACF i PACF), natomiast w rzeczywistości analizie poddane zostają wartości próbkowe. To powoduje, że charakterystyczne wartości dla modelu mogą być podane jedynie w przybliżeniu.

Może się również zdarzyć, że dla danych, które posiadamy, żaden model nie będzie odpowiedni. Nie jest wykluczone, że znajdzie się nie jeden, a kilka modeli, które będą się nadawały. Wówczas nie należy zapominać, że ostateczny dobór powinna poprzedzić weryfikacja poprawności dopasowania. W przypadku prognozowania szeregu czasowego powinno się zastosować także ocenę oraz porównanie skuteczności przygotowanych prognoz. Stosowanie reguł identyfikacji, które zostaną przedstawione w pracy ma sens tylko wtedy, gdy mamy do czynienia z szeregiem stacjonarnym. Jeżeli szereg jest niestacjonarny należy przygotować dane przed identyfikacją poprzez ich przekształcenie. Wśród podstawowych przekształceń można wymienić m.in. eliminację trendu i sezonowości poprzez różnicowanie lub na podstawie modelu dekompozycji, które zostaną użyte w praktycznej części pracy.

Kolejna reguła to identyfikacja modelu $WN(0, \sigma^2) \sim Z_t$ i $MA(q)$ na podstawie funkcji ACF. W tym miejscu należy wyjaśnić pojęcie białego szumu (ang. white noise) $WN(0, \sigma^2)$. Jest to ciąg nieskorelowanych zmiennych o jednakowym rozkładzie, dla którego wartość średnia równa się 0, a wariancja σ^2– funkcja teoretyczna autokorelacji jest postaci

$$ACF(h) = \begin{cases} 1 \; dla \; h = 0 \\ 0 \; dla \; h \neq 0 \end{cases}$$

Jeżeli autokorelacja próbkowa ACF(h) znajduje się pomiędzy przedziałami ufności dla h>q to spodziewamy się że dane są realizacją procesu średniej ruchomej.

2.9 Kryteria oceniające optymalizacje dopasowania AIC (Akaike Information Criterion)

Zwiększając liczbę parametrów możemy otrzymać lepsze dopasowanie modelu do danych, ale zbyt duże dopasowanie prowadzi do złych prognoz. Z tego powodu w postaci kryteriów wybory modeli istnieje składnik kary za wymiar modelu, tzn w ogólności mamy wzór

$$C(\text{model}) = -2\ln(L) + \text{kara}(\text{liczba współczynników modelu}),$$

gdzie L oznacza funkcję wiarygodności (likelichood)

Kryteria informatyczne różnią się postacią składnika kary. Tak dla AIC wzór wygląda tak: $AIC(p,q) = -2\ln(L)+2(p+q+k+1)$

Kryterium informacyjny Akaikiego jest jednym z najbardziej popularnych kryteriów wyboru pomiędzy modelami o różnej liczbie parametrów. Przy wyborze optymalnego modelu kierujemy się minimalizacją kryterium informacyjnego, wybieramy ten, dla którego wartość jest najmniejsza.

3. Analiza danych

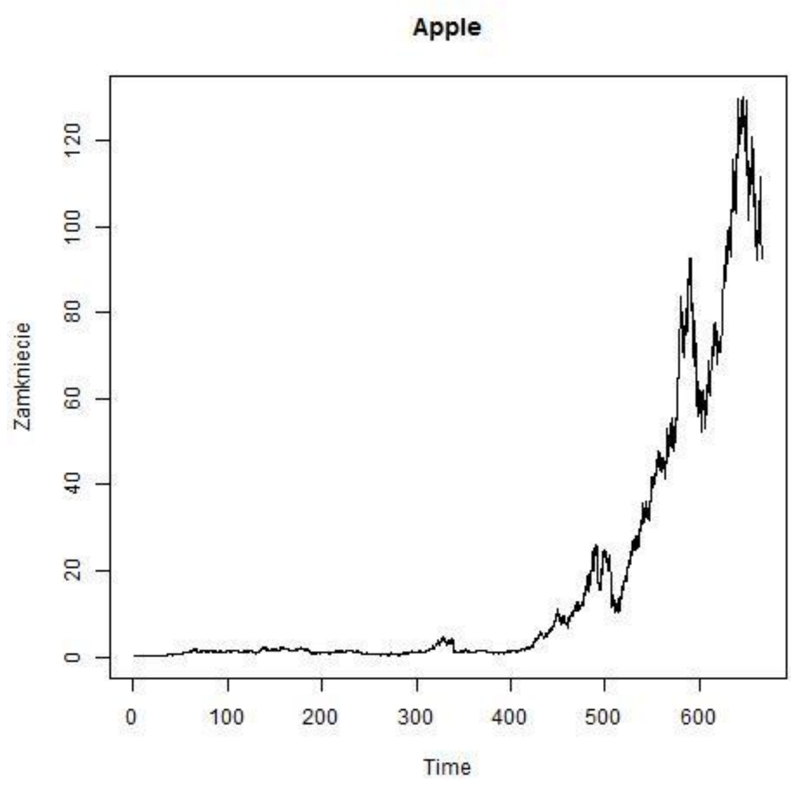
Analiza szeregu czsowego jest procesem wieloetapowym. Większość metod i modeli stosowanych w analizie szeregów wymaga wybrania adekwatnych parametrów.

Pierwszym etapem jest przedstawienie danych na wykresie i podstawowa analiza własności szeregu (obserwacja trendu oraz sezonowości). Żeby lepiej zrozumieć szereg analizowany pomocne może być wyodrębnienie występujących składowych: trend i sezonowość. Takie postępowanie nazywamy dekompozycją szeregu czasowego, które pomoże nam znależć odpowiedni model lub skonstuować dokładnieszy prognozy.

Drugi etap to przekształcenie danych (sprowadzenie szeregu do postaci stacjonarnej). Jedną z podstawowych metod jest różnicowanie. Transformacje oceniamy na podstawie funkcji ACF i PACF.

Trzeci etap - jeżeli po wyeliminowaniu trendu otrzymaliśmy szereg stacjonarny to możemy dopasować odpowiednie modele AR, MA, ARMA, ARIMA. Nailepszym modelem jest ten dla którego wartość AIC jest najmniejsza.

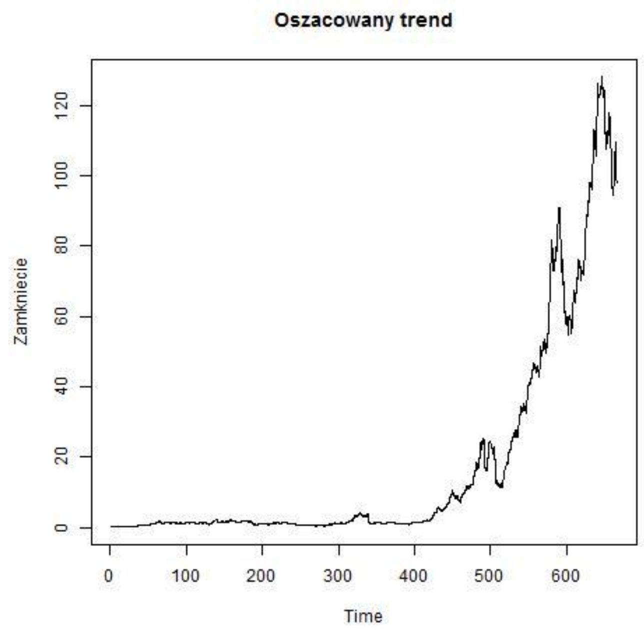
3.1 Analiza danych Apple



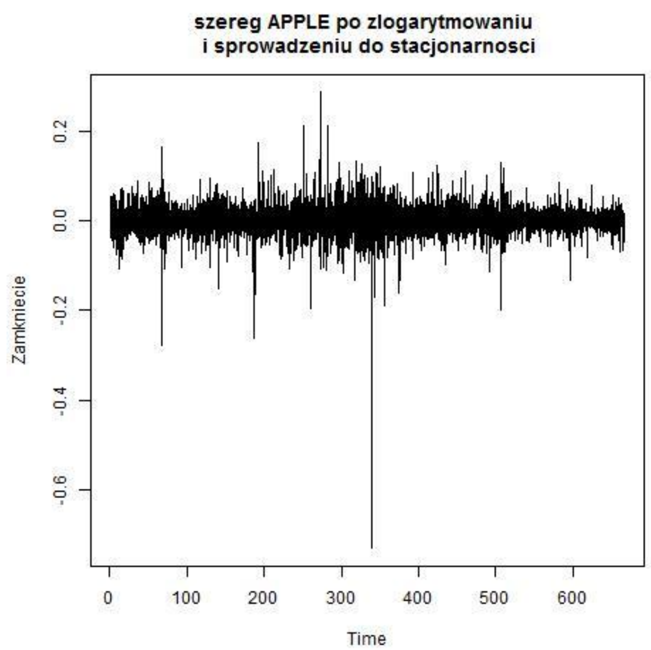
Na pierwszym etapie w analizie danych zaczynamy od ich wizualizacji. Szereg jest uznany za niestacjonarny ponieważ możemy zauważyć trend wykładniczy, co pokazuje, że wartość spółki rosła z czasem. Równanie naszego szeregu ma postać $X_t = m_t + Y_t$. Musimy sprowadzić go do postaci stacjonarnej. Analizowany szereg nie zawiera sezonowości, więc na początku musimy oszacować trend.

Rys.3.1 Miesięczne notowanie spółki Apple w okresie: 07.09. 1986- 06.05.2016.

Po oszacowaniu trendu, metodą wygładzania przy pomocy średniej kroczącej, która polega na zastąpieniu każdego elementu szeregu przez ważoną średnią n sąsiadujących wartości, gdzie n jest szerekością okna wygładzania, musimy go wyeliminować. Czyli nasze równanie wygląda tak: $m_t = X_t - Y_{t\,.}$ (rys 3.1.2) . Trend pokazuje nam wzrosty i spadki wartości spółki. Następnie uzyskujemy szereg reszt po usunięciu trendu (rys 3.1.3). Trend został wyeliminowany metodą różnicowania niesezonowego pierwszego rzędu (opóźnienie=1).



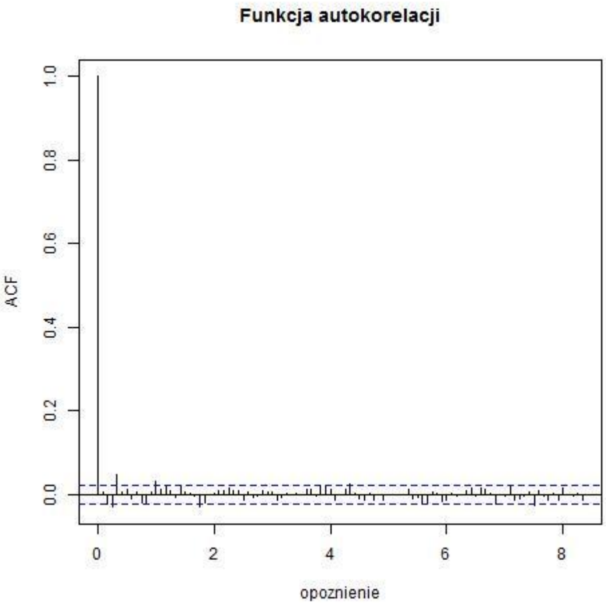
Rys.3.1.2 Oszacowany trend szeregu Apple z wykorzystaniem średniej ruchomej

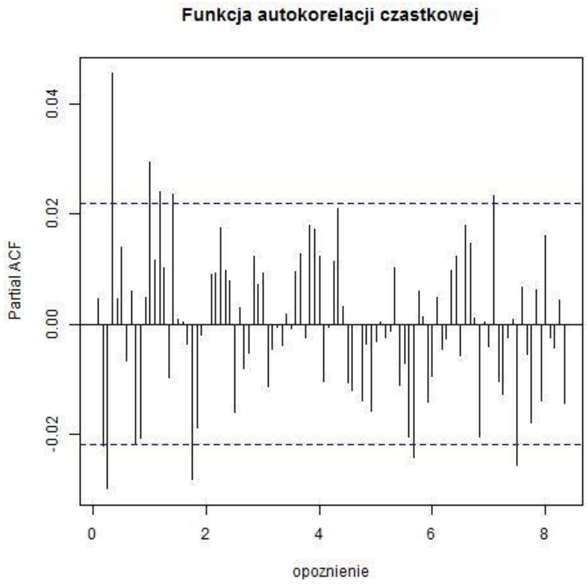


Rys.3.1.3 Szereg Apple po usunięciu trendu i sezonowości

Na rysunku 3.1.3 widzimy że po zastosowaniu różnicowania z opóźnieniem 1 trend zostal wyeliminowany. Otrzymaliśmy zatem szereg stacjonarny, co potwierdzają szybko zanikające funkcje ACF i PACF.

Kolejnym etapem analizy danych jest poszukiwanie parametrów p,q. Do tego użyteczna jest funkcja korelacji ACF i korelacji cząstkowej PACF. Kolegarmy funkcji autokorelacji pokazują że musimy uwzględnić parametr średniej ruchomej (q=3).



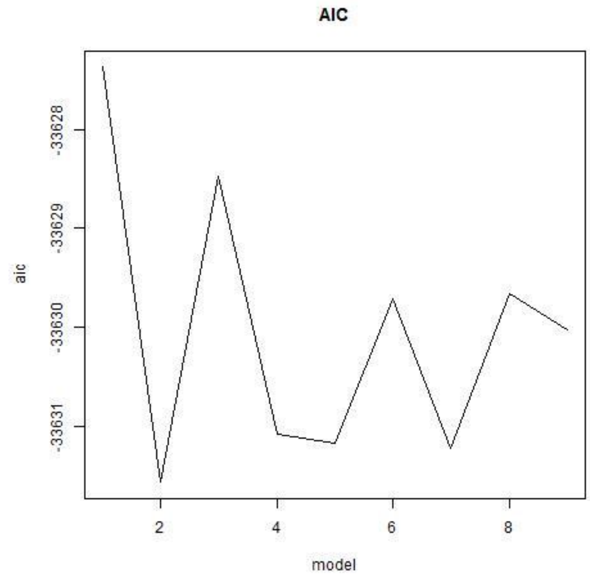


Rys.3.1.3 Funkcja ACF i PACF reszt szeregu Apple

Kolejnym krokiem jest budowanie modelu ARIMA. Najmniejsza wartość parametru AIC będzie naszym wyborem.

Tabela 1. Wartości AIC i Likelichood uzyskane po dopasowaniu modelu ARIMA szeregu Apple

		AIC	Likelichood
1	ARMA (3,3)	-33627.37	16821.69
2	**ARMA (3,4)**	**-33631.57**	**16824.78**
3	ARMA (3,5)	-33628.48	16824.24
4	ARMA (4,3)	-33631.09	16824.54
5	ARMA (4,4)	-33631.18	16825.59
6	ARMA (4,5)	-33629.72	16825.86
7	ARMA (5,3)	-33631.23	16825.61
8	ARMA (5,4)	-33629.66	16825.83
9	ARMA (5,5)	-33630.03	16827.01



Rys.3.1.4 Uzyskane wartości AIC dla różnych modeli ARM

Rysunek 3.1.4 pokazuje, że optymalnym jest model pod numerem 2 według wartości AIC. Porównując konkretne mode wybieramy ten gdzie wartość AIC jest najmniesza. Wybrany model - ARIMA(3,0,4).

Coefficients:

	ar1	ar2	ar3	ma1	ma2	ma3	ma4	intercept
	0.1936	0.2164	-0.4058	-0.1884	-0.2388	0.4257	0.0565	7e-04
s.e	0.1682	0.1285	0.1895	0.1680	0.1307	0.1908	0.0127	3e-04

sigma^2 estimated as 0.0008634: log likelihood = 16824.78, aic = -33631.57

Zatem szacowany model to:

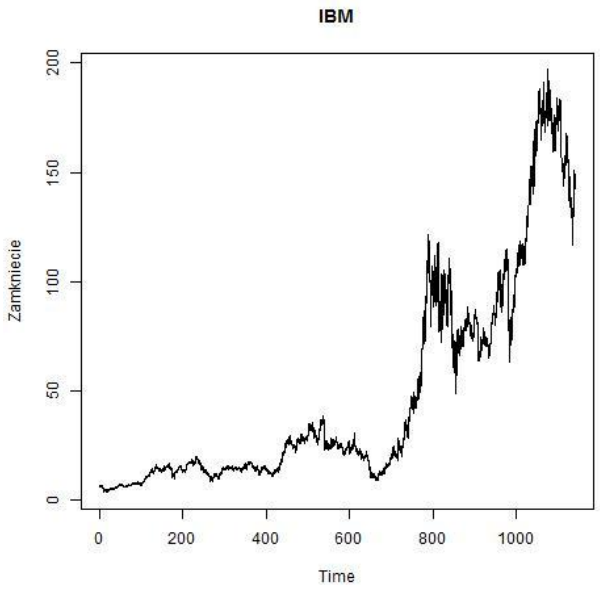
$$X_t = 0.194X_{t-1} + 0.216X_{t-2} - 0.451X_{t-3} + Z_t - 0.188Z_{t-1} - 0.239Z_{t-2} + 0.426Z_{t-3} + 0.057Z_{t-4}$$

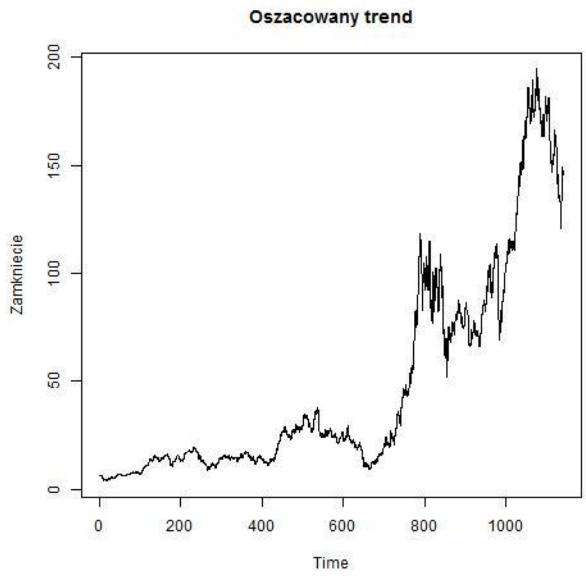
Model AR ma postać: $0.194X_{t-1} + 0.216X_{t-2} - 0.451X_{t-3}$

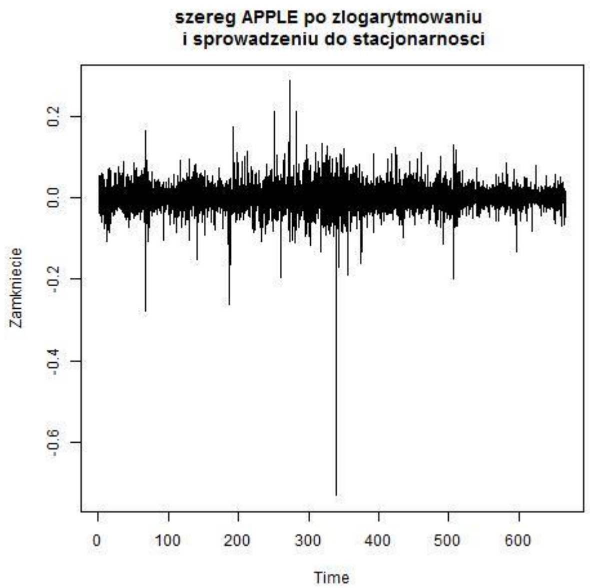
Model MA ma postać: $-0.188Z_{t-1} - 0.239Z_{t-2} + 0.426Z_{t-3} + 0.057Z_{t-4}$

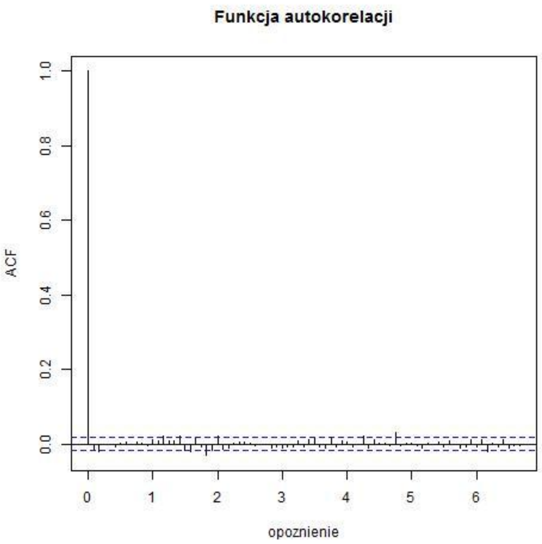
3.2 Analiza danych IBM

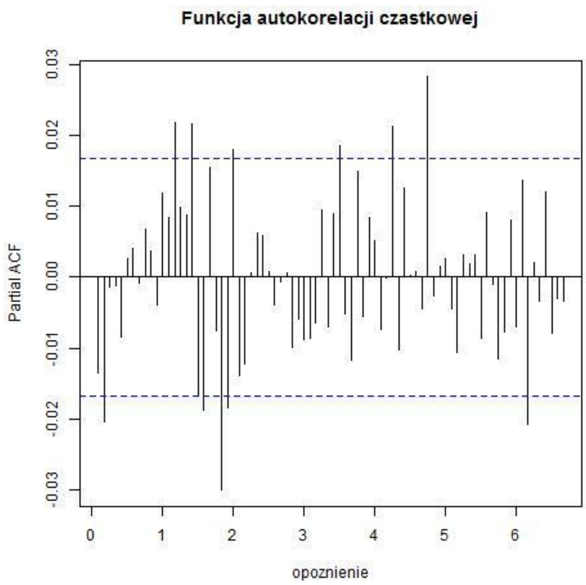
Analizę rozpoczyniemy od wizualizacji danych spółki IBM w okresie: 02.01. 1962 - 06.05.2016. Szereg również jest uznany za niestacjonarny rysunek prawy górny. Nie wykazuje sezonowości. Możemy zauważyć trend długoterminowy. Tak jak szereg wykazuje takie same wasności, co i szereg analizowany poprzednio, to będziemy wykonywać te same kroki.


IBM
Zamkniecie
Time
0
50
100
150
200
0
200
400
600
800
1000


Oszacowany trend
Zamkniecie
Time
0
50
100
150
200
0
200
400
600
800
1000


szereg APPLE po zlogarytmowaniu
i sprowadzeniu do stacjonarnosci
Zamkniecie
Time
0.2
0.0
-0.2
-0.4
-0.6
0
100
200
300
400
500
600


Funkcja autokorelacji
ACF
opoznienie
0.0
0.2
0.4
0.6
0.8
1.0
0
1
2
3
4
5
6


Funkcja autokorelacji czastkowej
Partial ACF
opoznienie
0.03
0.02
0.01
0.00
-0.01
-0.02
-0.03
0
1
2
3
4
5
6

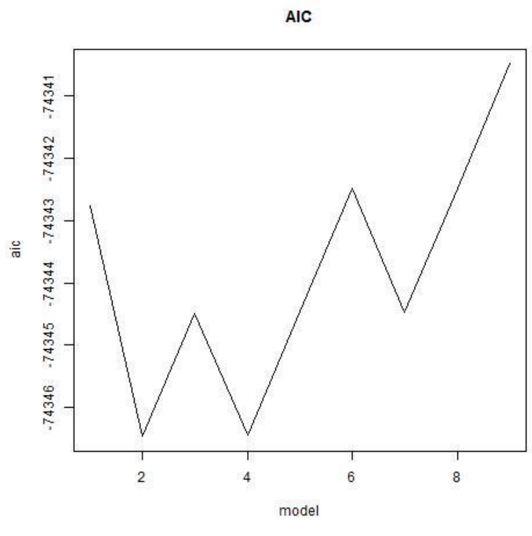

AIC
aic
model
-74341
-74342
-74343
-74344
-74345
-74346
2
4
6
8

Tabela 2. Wartości uzyskane po dopasowaniu modelu ARIMA szeregu IBM

		AIC	Likelihood
1	ARMA(1,1)	-74342.77	37175.39
2	**ARMA(1,2)**	**-74346.47**	**37178.24**
3	ARMA(1,3)	-74344.49	37178.25
4	**ARMA(2,1)**	**-74346.45**	**37178.22**
5	ARMA(2,2)	-74344.46	37178.23
6	ARMA(2,3)	-74342.49	37178.24
7	ARMA(3,1)	-74344.47	37178.24
8	ARMA(3,2)	-74342.48	37178.24
9	ARMA(3,3)	-74340.48	37178.24

Coefficients:

	ar1	ma1	ma2	intercept
	-0.0069	0.0069	-0.0204	2e-04
s.e	0.4363	0.4362	0.0101	1e-04

sigma^2 estimated as 0.0002547: log likelihood = 37178.24, aic = -74346.47

Coefficients:

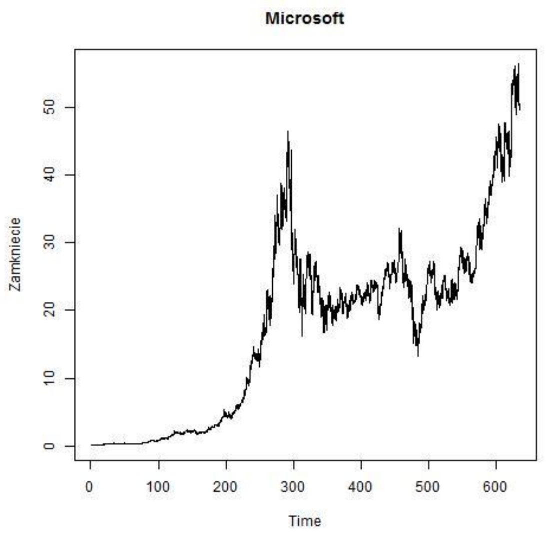
	ar1	ar2	ma1	intercept
	-0.0067	-0.0204	-0.0071	2e-04
s.e	0.4499	0.0102	0.4500	1e-04

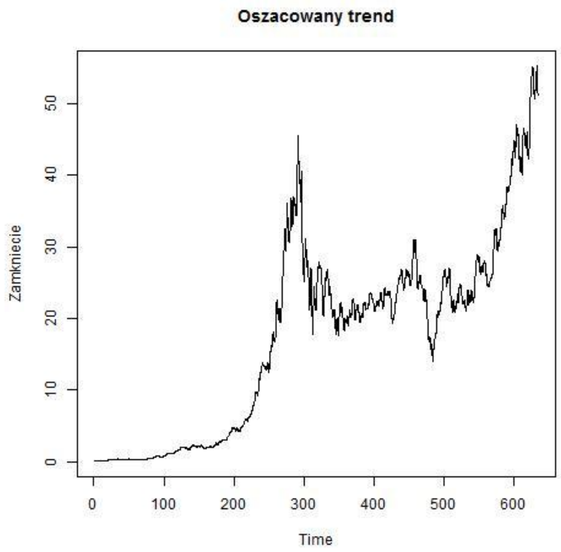
sigma^2 estimated as 0.0002547: log likelihood = 37178.22, aic = -74346.45

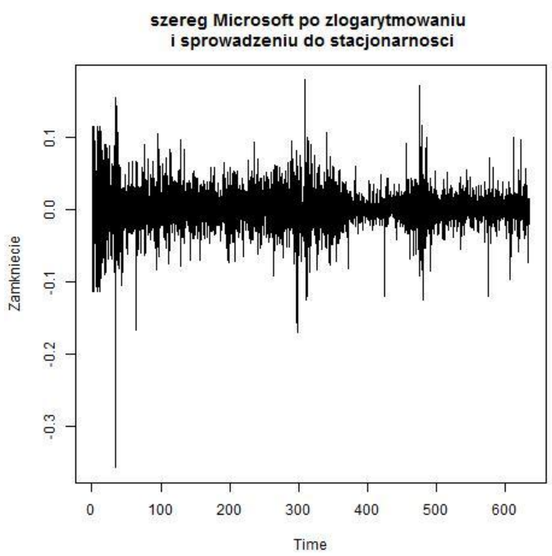
Na prawym górnym rysuneku widzimy oszacowany trend metodą średniej ruchomej. Lewy środkowy rysunek pokazuje nam szereg po logarytmowaniu i sprowadzeniu go do postaci stacjonarnej, metodą róznicowania pojedynczego. Teraz za pomocą funkcji autokorelacji, rysunek prawy środkowy, widzimy że szereg jest stacjonarny. O tym mówi szybko zanikająca funkcja ACF. Dany wykres pokazuje powinniśmy uwzględnić parametr q=1. Zachowanie funkcji cząstkowej autokorelacji PACF opisuje dolny lewy rysunek. Prawy dolny rysunek pokazuje że są dwa optymalnych modeli według otrzymanych wartości AIC jest ARIMA(1,0,2) i ARIMA(2,0,1)

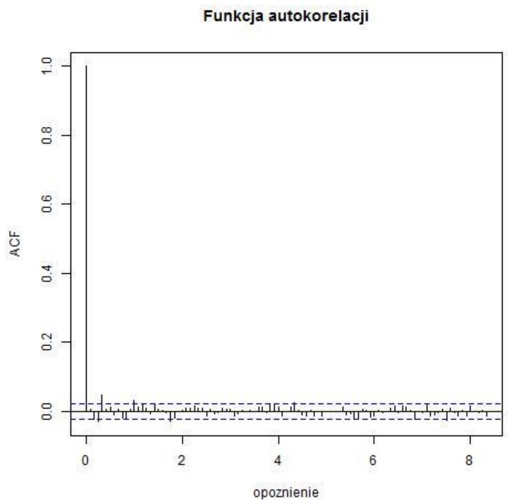
$$X_t = -0.007X_{t-1} - 0.02X_{t-2} + Z_t + -0.007Z_{t-1}$$

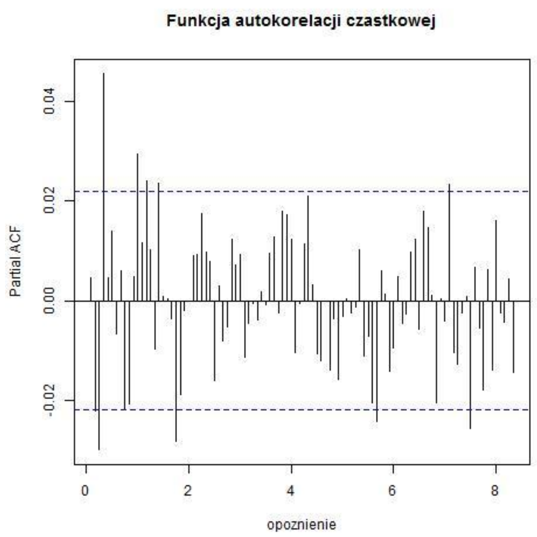
3.3 Analiza danych Microsoft











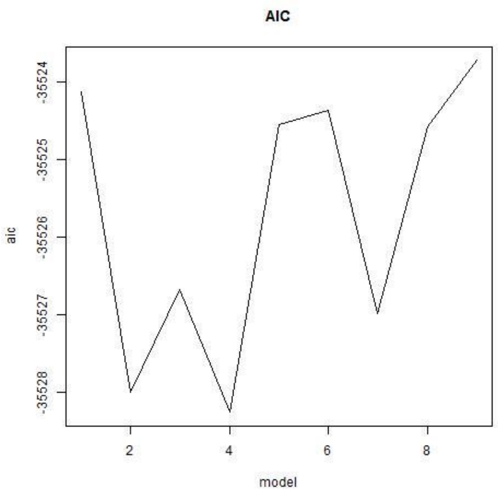


Tabela 3. Wartości AIC iuzyskane po dopasowaniu modelu ARIMA szeregu Microsoft

		AIC	Likelichood
1	ARMA(1,1)	-35524.13	17766.06
2	ARMA(1,2)	-35528.01	17769
3	ARMA(1,3)	-35526.68	17769.34
4	**ARMA(2,1)**	**-35528.26**	**17769.13**
5	ARMA(2,2)	-35524.55	17768.27
6	ARMA(2,3)	-35524.37	17769.19
7	ARMA(3,1)	-35526.99	17769.5
8	ARMA(3,2)	-35524.57	17769.29
9	ARMA(3,3)	-35523.72	17769.86

Coefficients:

	Ar1	Ar2	Ma1	intercept
	0.4858	-0.0361	-0.5264	9e-04
s.e	0.1054	0.0140	0.1051	2e-04

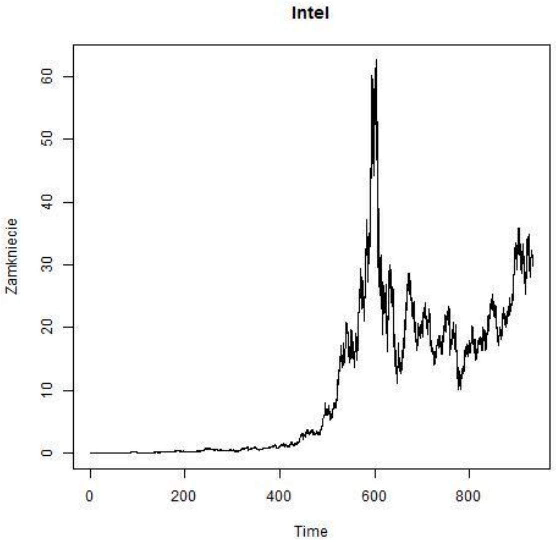
sigma^2 estimated as 0.0005451: log likelihood = 17769.13, aic = -35528.26

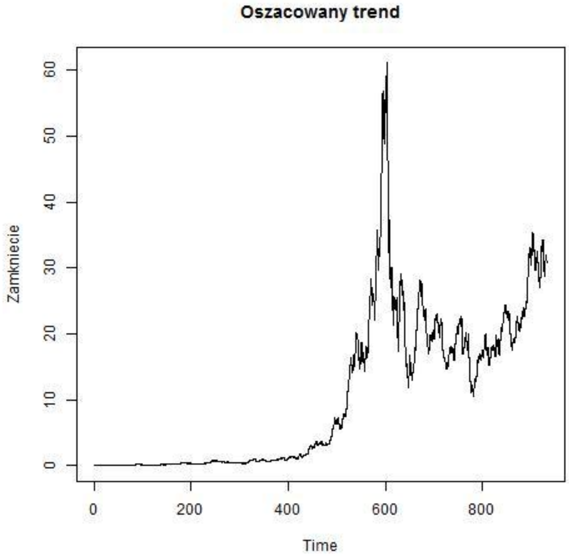
Zatem szacowany model to:

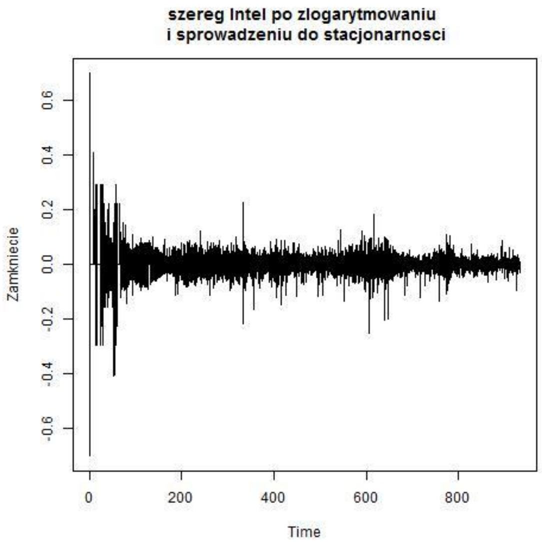
$$X_t = 0.486X_{t-1} - 0.036X_{t-2} + Z_t + -0.527Z_{t-1}$$

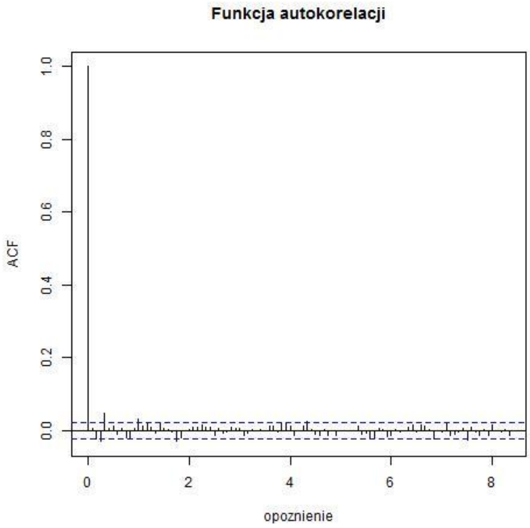
Lewy górny wykres zawiera notowanie misięczne spółki Microsoft w okresie: 13.03. 1986 - 06.05.2016. Tak jak szereg uznany za niestacjonarny, możemy zauważyć trend wykładniczy. Szereg również nie zawiera sezonowości. Musimy sprowadzić go do postaci stacjonarnej, czyli wyeliminować trend. Na prawym górnym rysuneku widzimy oszacowany trend metodą średniej ruchomej. Lewy środkowy rysunek pokazuje nam szereg po logarytmowaniu i sprowadzeniu go do postaci stacjonarnej, metodą róznicowania pojedynczego. Teraz za pomocą funkcji autokorelscji, rysunek prawy środkowy, widzimy że szereg jest stacjonarny. O tym mówi szybko zanikająca funkcja ACF. Dany wykres pokazyje powinniśmy uwzględnić parametr q=2. Zachowanie funkcji cząstkowej autokorelacji PACF opisuje dolny lewy rysunek. Prawy dolny rysunek pokazuje że optymalnym modelem według otrzymanych wartości AIC jest ARIMA(2,0,1)

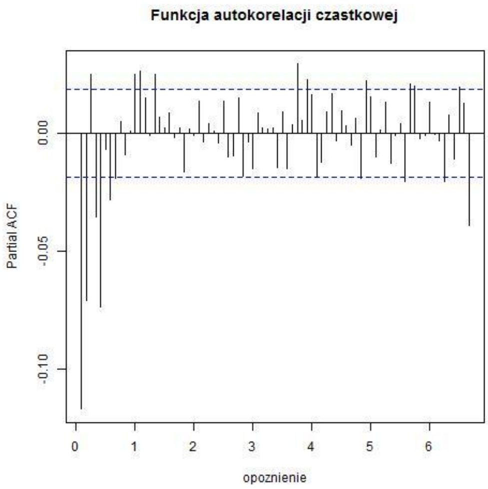
3.4 Analiza danych INTEL











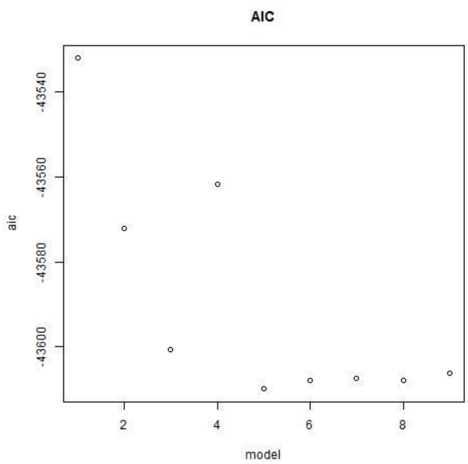


Tabela 4. Wartości AIC i Likelichood uzyskane po dopasowaniu modelu ARIMA szeregu INTEL

		AIC	Likelichood
1	ARMA(2,2)	-43532.14	21772.07
2	ARMA(2,3)	-43572.22	21793.11
3	ARMA(2,4)	-43600.56	21808.28
4	ARMA(3,2)	-43561.74	21787.87
5	**ARMA(3,3)**	**-43609.88**	**21812.94**
6	ARMA(3,4)	-43608.01	21813
7	ARMA(4,2)	-43607.32	21811.66
8	ARMA(4,3)	-43607.97	21812.99
9	ARMA(4,4)	-43606.1	21813.05

Coefficients:

	Ar1	Ar2	Ar3	Ma1	Ma2	Ma3	intercept
	0.0155	-0.2363	0.4858	-0.1470	0.1747	-0.4852	7e-04
s.e	0.0805	0.0671	0.0402	0.0818	0.0699	0.0458	2e-04

sigma^2 estimated as 0.001179: log likelihood = 21812.94, aic = -43609.88

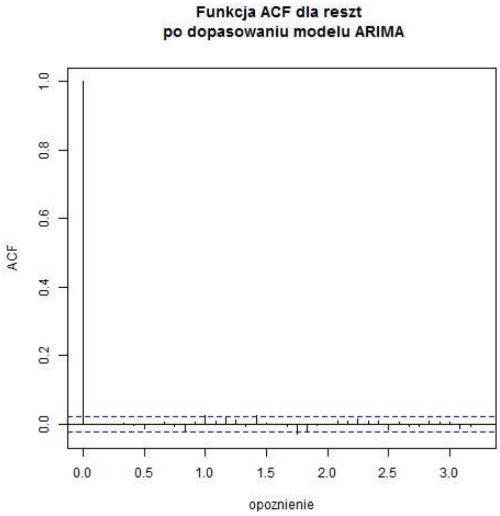
Szacowany model to:

$$X_t = 0.016X_{t-1} - 0.236X_{t-2} + 0.486X_{t-3} + Z_t - 0.147Z_{t-1} + 0.175Z_{t-2} - 0.485Z_{t-3}$$

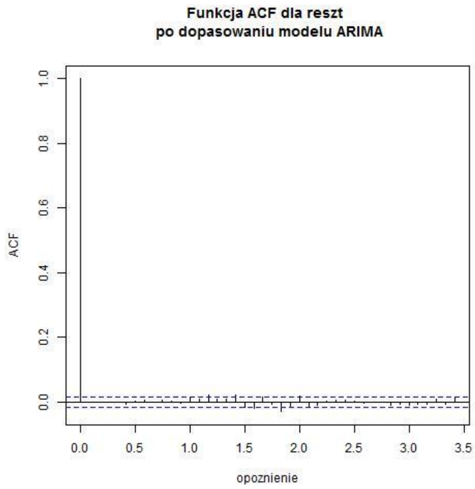
Lewy górny wykres zawiera notowanie misięczne spółki INTEL w okresie: 01.02. 1972 - 06.05.2016 . Analizowany szereg jest niestacjonarny, ponieważ możemy zauważyć trend wykładniczy. Szereg również nie wykazuje sezonowości. Dlatego sprowadzamy go do postaci stacjonarnej, czyli usuwamy trend. Na prawym górnym rysunku widzimy oszacowany trend metodą średniej kroczącej. Lewy środkowy rysunek pokazuje nam szereg po logarytmowaniu i sprowadzeniu go do postaci stacjonarnej, metodą różnicowania pojedynczego. Dalej za pomocą funkcji autokorelacji, rysunek prawy środkowy, widzimy że szereg jest stacjonarny. O tym mówi szybko zanikająca funkcja ACF. Dany wykres pokazuje powinniśmy uwzględnić parametr q=2. Zachowanie funkcji cząstkowej autokorelacji PACF opisuje dolny lewy rysunek. Prawy dolny rysunek pokazuje że, optymalnym modelem według otrzymanych wartości AIC jest ARIMA(3,0,3)

Wnioski

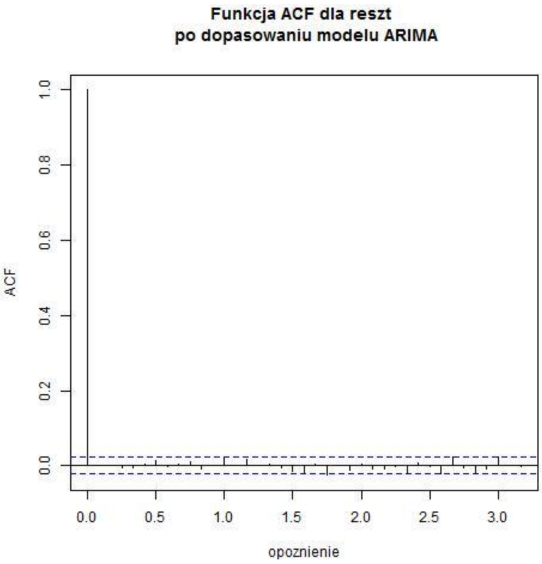
Celem niniejszej pracy była analiza techniczna szeregów czsowych na przykładzie spólek IBM, Microsoft, Apple, Intel. W pracy zostalo zareprezentowano zastosowanie metody ARIMA służącej do analizy szeregu czasowego z trendem. Wyniki obliczeń z wykorzystaniem modelu ARIMA przedstawione korzystając z pakietu R. Analizując przykłady można było zauważyć wspólne cechy takie jak: żaden z przykładów nie wykazywał sezonowości, a tylko zawierał trend wzrostowy. Dla spółek Microsoft, IBM optymalne modele są jednakowe. Analizowane szeregi miały jednakowe własnosci, czyli analiza postępowania była taka sama.



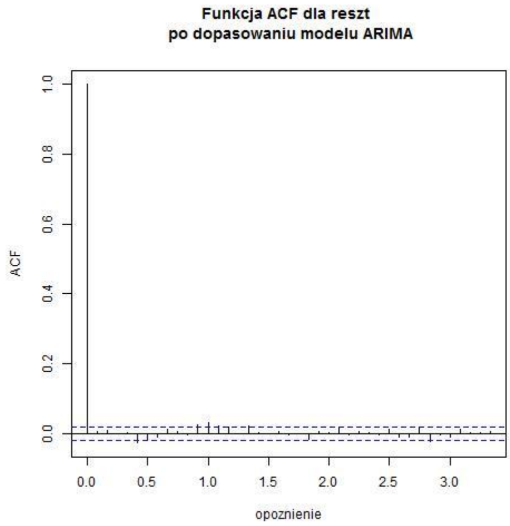
a) Apple



b) IBM



c)Microsoft



d)INTEL

Na podstawie uzyskanych kolegramów funkcji ACF możemy stwierdzić że wybrane modeli jest poprawne do opisu szeregów Apple, IBM, Microsoft, INTEL.Rozkład reszt nie odbiega od rozkładu normalnego, co pozwala przyjąć, że reszty są procesem białego szumu.

Bibliografia

[1] J.Adler, R in a Nutshell. A Desktop Quick Reference, O'Reilly Media, 2012

[2] A.Zagdański, A. Suchwałko „Analiza i prognozowanie szeregów czasowych praktyczne wprowadzenie na podstawie środowiska R", Warszawa 2016.

[3] P.J. Brockwell and R.A.Davis, Time Series Analysis, Forecasting and Control, Holden- Day, Incorporated, 1994.

[4] P.J. Brockwell and R.A.Davis, Time Series: theory and methods, Springer-Verlag New York, Inc., New York, NY, USA, 2ed edition, 1991.

[5] Brooks Ch., 2002: Introductory econometrics for finance, Cambridge University Press, rozdział 5.

[6] Cieślak M., 1996: Prognozowanie gospodarcze, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego, Wrocław,.

[7] Ekonometria. Metody, przykłady, zadania, 2002: (red.) J. Dziechciarz, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego, Wrocław.

[8] Gajda J., 2001: Prognozowanie i symulacje a decyzje gospodarcze, Wydawnictwo C.H. Beck, Warszawa.

[9] M. Gągolewski, Programowanie w języku R. Analiza danych. Obliczenia. Symulacje. PWN, 2014

[10] R. J. Hyndman and Y. Khandakar, Automatic time series forecasting: The forecast package for R, Journal of Statistical Software, 27(3):1-22, March 2008

[11] Maddala G. S., 2006: Ekonometria, PWN, Warszawa, rozdział 13.

[12] S.Łobejko, K. Masłowska, R.Wojdan , Analiza i prognozowanie szeregów czsowych я programem SAS, Warszawa 2014.